AWARD NUMBER: **W81XWH-14-1-0472**

TITLE: Oncogenic LINE-1 Retroelements Sustain Prostate Tumor Cells and Promote Metastatic Progression

PRINCIPAL INVESTIGATOR: **Denise S. O'Keefe, Ph.D.**

CONTRACTING ORGANIZATION:

UNIVERSITY OF TEXAS , San Antonio
SAN ANTONIO TX 78229-3901

REPORT DATE: December 2016

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2016	Final	30-Sep-2014 to 29-Sep-2016

4. TITLE AND SUBTITLE
Oncogenic LINE-1 Retroelements Sustain Prostate Tumor Cells and Promote Metastatic Progression

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-14-1-0472

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Denise S. O'Keefe, Ph.D., Dean Bacich, Ph.D., Shahida Flores, M.S., Wasim Chowdhury, M.S., Ping Wu, Ph.D. and Yidong Chen, Ph.D.

E-Mail: okeefed@uthscsa.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

UNIVERSITY OF TEXAS, San Antonio
OFFICE OF GRANTS MANAGEMENT
7703 FLOYD CURL DR
SAN ANTONIO TX 78229-3901

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this hypothesis development project was to determine if expression of LINE-1 elements in prostate tumor metastases contribute to its progression by activating oncogenic DNA sequences, or silencing tumor suppressor like sequences. We have RNA-sequencing data that we developed novel pipelines to analyze what is typically called "junk sequence" and removed from standard RNA-seq analysis pipelines, and have developed a database of novel sequences that are expressed in lymph node metastases from prostate cancer, and contain a portion of LINE-1 element, in addition to a portion of another transcript. Interestingly, the standard analysis pipeline suggests significant activation of non-coding RNA sequences as well. Furthermore, we cloned a repressor of LINE-1 retroelements, the PIWIL1 gene, it put it under the control of a doxycycline-inducible promoter. Expression of this in LNCaP and PC-3 prostate cancer cells was robust, but had no effect on the cells; as long as three single-nucleotide variations were present. When wild-type PIWIL1 was expressed induction was nearly impossible suggesting that the agonaute interacting domain is critical to the function of this protein and cells expressing it can not live, potentially providing proof-of-principle for future gene-based therapeutics for this cancer.

15. SUBJECT TERMS
Prostate cancer, LINE-1, PIWIL1, retrotransposons

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	18	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. **Introduction…….4**
2. **Keywords…….4**
3. **Accomplishments……..……..4**
4. **Impact……...……6**
5. **Changes/Problems...…..……7**
6. **Products…….……..…..…….8**
7. **Participants & Other Collaborating Organizations……8**
8. **Special Reporting Requirements……8**
9. **Appendices……10**

1. INTRODUCTION

Prostate tumors are generally under-methylated at a global level; we have shown that the most de-methylated tumors are in fact those metastatic to lymph nodes, when compared to the primary tumor or normal prostate adjacent to tumors. The function of global DNA methylation at repetitive sequences such as LINE-1 (sometimes called L1s) in normal cells, is to suppress activation of these ancient retro-virus like sequences, which once located in the human genome are referred to as retrotransposons. DNA methylation of these sequences represses their activation, and conversely, demethylation at these sequences activates LINE-1 transcription. While only 80-100 LINE-1 elements in the human genome are considered capable of retrotranspostion resulting in their replication and re-location elsewhere in the genome, there are approximately 500,000 LINE-1 elements, comprising 17% of total genome. While many of these sequences are mutated many are still capable of transcription. For this Hypothesis Development proposal, we postulated that the demethylation of LINE-1 sequences we had observed in prostate tumors would result in activation of these sequences, and that this was one potential mechanism by which cancer progression could be mediated. Furthermore, we hypothesized that PIWIL1, a protein which is normally expressed in testis where it is known to repress LINE-1 expression that occurs during spermatogenesis, could block LINE-1 transcriptional activation and therefore the growth of prostate cancer metastases and provide proof-of-principle that we would have a novel therapeutic target present in virtually all prostate cancer metastases.

2. KEYWORDS

Piwil1 retrotransposon LINE-1 agonaute prostate

3. ACCOMPLISHMENTS

Major goals of the project

The project was divided into three main tasks; for Task 1, the goal was to determine whether expression of LINE-1 encoded proteins can transform non-tumorigenic prostate cells and whether LINE-1 ORF expression confers a metastatic phenotype. For Task 2, the plan was to determine the effect active LINE-1 transcription has on gene expression in metastatic prca. And for Task 3, we planned to provide “proof of principle” for a potential therapy targeted at LINE-1 expression in advanced prostate cancer. Task 4 was writing up the results and presenting at a conference.

Accomplishments related to goals

This project was funded by a high-risk/ high reward mechanism, the hypothesis development mechanism. According to the summary statement received at the review, the project was over-ambitious but should yield some interesting data. This was correct, this was a very difficult project. The major findings, described in more detail in the appendix would be two-fold; firstly, that PIWIL1, the protein which we hypothesized would kill prostate cancer cells when expressed

appears to be correct; when we first induced expression of this transcript in the LNCaP and PC3 prostate cancer cell lines, there was robust (>200-fold) induction – yet no protein expression at all. Thorough sequence analysis revealed that the coding sequence was in-frame, however when compared to an updated reference sequence at NCBI for PIWIL1, there were three differences resulting in three amino acid changes in the PIWIL1 protein. One of these, found in the PIWI-interaction domain (agonaute-like), was altered to a proline, likely disrupting both the activity and the structure of the protein and thereby we would predict resulting in its untimely degradation. We synthesized and re-subcloned the new version of the sequence into the same vector, however we were unable to induce this to more than 40-fold approximately in LNCaP cells, despite using the same conditions that resulted in robust transcript induction previously. In PC3 cells, we were able to get less than a 2-fold induction. No protein for PIWIL1 was detectable, except in positive controls (mouse testis). We think that this implies that PIWIL1 expression is so toxic to the tumor cells that even "leaky" expression in an inducible transcription system was able to result in the cell death of any cells expressing it, and in future a less inducible, but less leaky system should be used – for example the ecydsone system.

The other major goal of this project was to identify hybrid LINE-1 sequences activated by demethylation of the prostate tumor genome. LINE-1 transcription can theoretically extend into neighboring genes either producing sense of antisense transcripts resulting in novel gene activation or inactivation. This data is automatically masked from most RNA sequencing data, as repeat sequences are filtered out of the data. We developed a novel approach to identify and examine these sequences in the sequencing data from 4 independent lymph node metastases of prostate cancer. We determined that there are in fact common transcripts among all and / or some of the samples that consist of LINE-1 sequences on one part of the transcript, and a particular gene that might be involved in the metastatic potential of the cell. The top three candidates we identified were the LPP, AAK1 and WDR72 genes. These genes have not been studied in much detail at all, and none had been implicated in prostate cancer before this study. Developing the pipeline for this study was difficult, and can be improved however, which may make it more sensitive. We are continuing the study in our lab however this is the final report for this project. Details of the methods and some of the more interesting results of these studies can be found in the Appendix of this document.

Opportunities for training and professional development provided by the project

While training was not a goal of this project, a graduate student worked one-on-one beside myself while developing the constructs and cell lines for the project, and determining PIWIL1 expression in them. In addition, I had a summer student work on the project – he had recently completed his master's degree in biology/bioinformatics, and wanted to learn about cancer biology. I worked side-by-side with him to teach him how to use next-generation sequencing analysis tools to analyze human sequencing data. He also prepared patient samples for generation of the sequencing libraries, and plans to continue working in his spare time on this project starting this summer (although this is a final report for this project, we have generated a lot of data that can still be analyzed further). Finally, I myself took online courses in bioinformatics and next-generation sequencing analysis to help understand how to analyze the data for this project when it became clear that there was no standard analysis pipeline for this kind of data. The courses I took were during weekends and after hours and weren't funded by this project, but doing this project provided the opportunity to learn how to analyze this data. Therefore the project has served to increase the professional development of graduate students in the areas of prostate cancer research and molecular biology techniques (cloning, sequence analysis, lentiviral construction and cell transduction, gene expression, western blotting and

protein expression, human sequence analysis), and my own professional development with regards to knowledge in the area of bioinformatics.

Dissemination of results to communities of interest

I am a co-investigator on a DOD-funded HBCU summer training program, which recruits undergraduate students from Huston-Tillotson University in Austin, Texas and provides a summer research experience for them where they learn about prostate cancer research, as well as get to carry out a research project of their own. As part of this experience I give a lecture to the students and also talk about our research and why we are doing each project in the laboratory, there are four students in this program each year and I interacted with them both one-on-one and as a group. For all of these students, this was their first exposure to medical research, which encompasses prostate cancer research. The goal was to increase their interest in pursuing a career in prostate cancer research, and also to inform them about the clinical aspects of prostate cancer. Furthermore, we plan to continue this project and publish on it.

Plans for the next reporting period to accomplish the goals

Nothing to report – final report

4. IMPACT

Impact on the development of the principal discipline of the project

We generated novel prostate cancer cell lines that are able to turn on a potentially therapeutic pathway when stimulated with the common antibiotic, tetracycline. Currently there is confusion in the literature regarding the role of this pathway, mediated by the gene PIWIL1, and whether it is pro- or anti- oncogenic. We have proposed that it is anti-oncogenic, and only one of the National Cancer Institute's panel of 60 cancer cell lines expresses any transcripts of this gene, and even then at 0.9 transcripts per million transcripts, also suggesting that it is not oncogenic and in fact may be a potent anti-cancer gene. When we used a common variant of the gene we were able to induce its transcription in both LNCaP and PC3 prostate cancer cell lines, however there was no effect on cell growth. On close re-examination of the sequence we realized that three bases were variable amongst people, they were polymorphic (single nucleotide polymorphisms). One of these mapped to the PIWI-interacting domain, which is likely to be very important in mediating the function of the protein. When we corrected the sequence for all 3 of these polymorphisms, we were no longer able to get inducible expression of the construct, strongly suggesting that the minor changes we had made had indeed had a major effect on the willingness of the cancer cells to allow its expression. This is likely due to some low level expression of the PIWI protein without induction leading to potent cell death of cells that express it, thus selecting for cells that were able to suppress its expression. We will be using this data to apply for NIH funding to examine this phenomenon further. Furthermore, we have developed a pipe-line for analyzing RNA-sequencing data that is typically thrown away by traditional analysis pipe-lines yet may yield important data with regards to regulation of gene expression in cancer.

Impact on other disciplines

The inducible PIWIL1 construct will be useful to people studying fields other than cancer, for example embryogenesis and reproduction, as PIWIL1 is important in these processes.

Impact on technology transfer

Nothing to report

Impact on society beyond science and technology

Nothing to report

5. CHANGES/PROBLEMS

Changes in approach and reasons for change

Nothing to report

Actual or anticipated problems or delays and plans to resolve them

This project had plenty of problems, from the fact that it began as I was setting up a new laboratory, had a graduate student leave and a research technician move on to medical school, leaving us short-handed. Additionally, in the first report we described the difficulties we have in obtaining lymph node metastases from patients as the banked tissues had degraded RNA. However, we resolved this by working with colleagues from pathology and the clinicians in our department so that new metastasis tissues were banked in a manner that facilitated maintenance of RNA integrity. Other problems were that the original PIWIL1 construct had three polymorphic sequences in it, and did not seem to produce protein however it did produce RNA transcripts. Finally the analysis of LINE-1 hybrid sequences was not a standard method and I had to work with a bioinformatician to help do an appropriate analysis. It is likely that the analysis could still be further refined as well, and we are working on this despite the fact that the project is finished.

Changes that had a significant impact on expenditures

Nothing to report

Significant changes in use or care of human subjects, vertebrate animals, biohazards and/or select agents

Nothing to report

Significant changes in use or care of human subjects

Nothing to report

Significant changes in use or care of vertebrate animals

Nothing to report

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS

Publications, conference papers and presentations

Nothing to report

Websites or internet sites

Nothing to report

Technologies or techniques

Nothing to report

Inventions, patent applications and/or licenses

Nothing to report

Other products

- We have generated paired-end RNA sequencing data, including analyses that identify novel expressed sequences from typically difficult to obtain lymph node metastases from carcinoma of the prostate (approximately 300 million reads)
- We have generated novel cell lines transduced with inducible PIWIL1 expression vectors carrying either the wild-type PIWIL1 sequence or a sequence with polymorphic sites (3) (two of which alter the predicted active enzymatic site of the protein).
-
- When we publish this data I believe that we will have significantly improved the understanding of the biological important of the PIWIL1 protein. The fact that it is suppressed by cancer cells when the PIWI-interacting domain is mutated, strongly suggests that this protein can in fact act as an inhibitor of tumor growth although unfortunately we will have to find another way to introduce it to cells to prove this conclusively. This would likely mean using another vector that produces less induction but that is tightly controlled (no leakage at all from the promoter).

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

Individuals that worked on this project for at least one calendar month in the final year

Name:	Denise O'Keefe
Project Role:	Principal Investigator
Researcher Identifier:	orcid.org/0000-0002-3992-437X
Nearest person month worked:	3

Contribution to Project:	Overseeing project/training workers in the techniques, bioinformatics analysis
Funding Support:	
Name:	Dean Bacich
Project Role:	Co-Investigator
Researcher Identifier:	orcid.org/0000-0002-1175-4241
Nearest person month worked:	1
Contribution to Project:	training workers in the techniques
Funding Support:	
Name:	Shahida Flores, M.S.
Project Role:	Graduate Student
Researcher Identifier:	orcid.org/0000-0002-3852-4307
Nearest person month worked:	7
Contribution to Project:	Cloning, development of the cell lines to express PIWIL1 (first construct)
Funding Support:	O'Keefe startup funds
Name:	Ping Wu, Ph.D.
Project Role:	Project scientist
Researcher Identifier:	Not known
Nearest person month worked:	4
Contribution to Project:	Cloning and development of cell lines (second PIWIL1 construct)
Funding Support:	Department of Urology
Name:	Hao Zhang, M.S.
Project Role:	Summer student
Researcher Identifier:	Not known
Nearest person month worked:	1
Contribution to Project:	Preparation of RNA from patient samples for RNA-sequencing; bioinformatics analysis
Funding Support:	NIH 5P20CA165589 THE CANCER BIOINFORMATICS INITIATIVE: A UTSA/UTHSCSA PARTNERSHIP

Changes in active other support of the PD/PI or senior/key personnel during the last reporting period

Nothing to report

Other Organizations involved as partners

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS

Nothing to report

Appendix

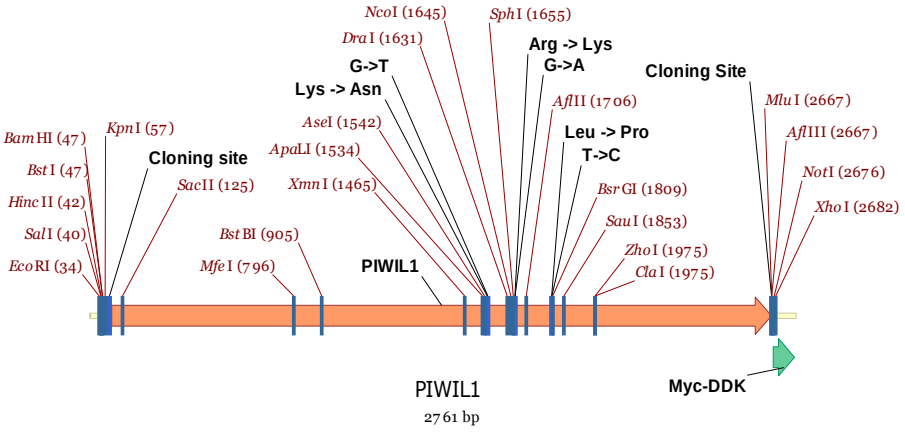
1. Generation of PIWIL1 inducible lentiviral constructs:

A TrueORF gold expression-validated cDNA clone was obtained from Origene (NM_004764). This construct contained the human piwi-like 1 gene transcript variant 1 in the pCMV6-Entry vector, with a c-terminal myc-DDK tag. The following primers were used to generate PCR product containing the PIWIL1 gene (without the myc tag in case it would interfere with function), utilizing high-fidelity taq polymerase and following instructions in the Lenti-X Tet-One inducible expression system (Clontech), to clone into the pLVX-tet-one vector. Forward primer: 5' ccc tcg taa aga att cat gac tgg gag agc ccg agc cag a 3' and the reverse primer was 5' gag gtg gtc tgg atc ctt aga ggt agt aaa ggc ggt ttg a 3' (see Figure 1.)

This was achieved and the plasmid completely sequenced to determine that it was identical to the original source from Origene, which it was. Lentivirus was produced by transfecting the plasmid into the HEK293T cell line (Clontech), and collecting infective virus. LNCaP and PC3 prostate cancer cells were transduced with the pLVX-tet-one-PIWIL1 virus and separately with a Green Fluorescent Protein-expressing construct as a control. Cells were then placed under selection with puromycin. After selection for several weeks, cells were treated with either zero, 100 or 1000ng of dox (a tetracycline analog), to induce transcription of PIWIL1. Prior to this all media had used fetal bovine serum that was devoid of tetracycline so as there could be no "leaky" expression of the construct, as if our hypothesis was correct and it was indeed a tumor-suppressor gene, this leaky expression might result in the death of any cells capable of expressing it without our realization.

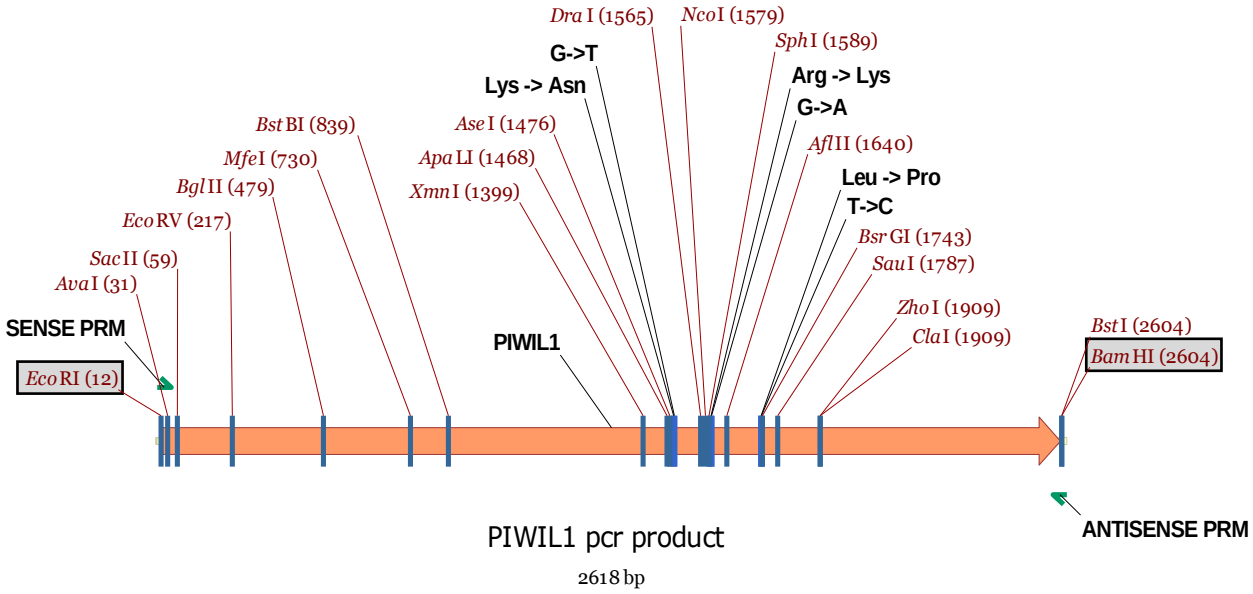
Figure 1. Inducible PIWIL1 cloning strategy.

A) PIWIL1 plasmid sequence from Origene (Cat# RC205469)



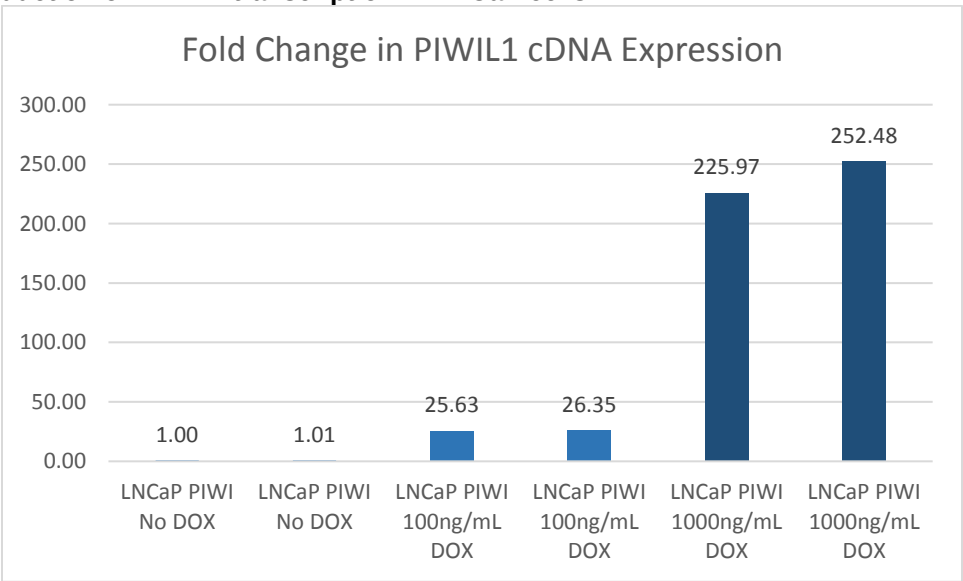
PCR amplified using the following two primers:
F PIWIL1 cloning: 5' CCCTCGTAAAGAATT-CATGACTGGGAGAGCCCGAGCCAGA 3'
R PIWIL1 cloning: 5' GAGGTGGTCTGGATCCTTA-GAGGTAGTAAAGGCGGTTTGA 3'

B) PIWIL1 PCR Product:



This PCR Product was cut with EcoRI & BamHI, the 2592 bp product was cloned into pLVX-TetOne-Puro vector using the same sites. The underlined sequence in the reverse primer introduces a stop codon right before the BamHI site.

Figure 2. Induction of PIWIL1 transcription in LNCaP cells.



Real-time PCR was used to assess expression of the PIWIL1 clones after induction (see Figure 2). Briefly, stable cell lines carrying the pLVX-TetOne-Puro-PIWIL1 vector were generated. Stable lines were then treated with no, 10, 100 or 1000ng/ml doxycycline, which is a tetracycline analog capable of activating the TetOne promoter. Quantitative analysis relative to GUSB (a housekeeper gene) showed that there was an average of over 200-fold increase in PIWI gene expression after induction with 1000ng/ml DOX, after 48 hours treatment. Next, we utilized antibodies against human

PIWI, to detect expression of the protein on a western blot. As one antibody anti-PIWIL1 PA5-19457 (Pierce) was generated against a peptide from 800-861 of the amino acid sequence of PIWIL1, which is 95% identical to mouse PIWIL1, we used a lysate of testis protein from mouse as a positive control for the blot (Figure 3). It should be noted that we tried other antibodies against PIWI, however were not able to detect proteins as there are no cell lines that express this endogenously at appreciable levels that we could use for a positive control.

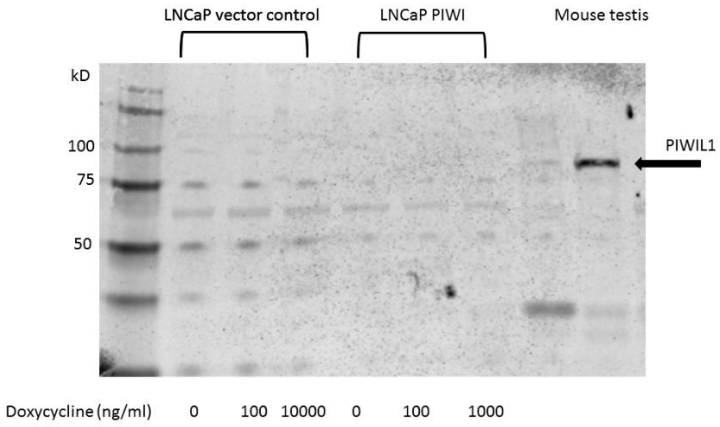


Figure 3. Western blot analysis of stable LNCaP-tetone-puro-PIWIL1 cell line induction. Mouse testis was used as a positive control to demonstrate PIWIL1 protein, however the LNCaP stable cell line that was able to induce PIWI mRNA expression as shown in Figure 2, does not appear to express PIWIL1 protein. Similar results were seen for the PC3 cell line, and additionally there was no cell death seen in either cell line upon transcriptional induction, which was contrary to our hypothesis.

At this point we returned to the sequence analysis; while our PIWIL1 cloning indeed produced an identical sequence to the original clone from Origene, there were three differences that we believed to be single-nucleotide polymorphisms (SNPs), as in natural variation. Compared to the updated PIWIL1 gene (NM_004764.4) sequence in the National Center for Biotechnology database there were three differences, which could be predicted to result in three changes to the protein sequence (Figure 4).

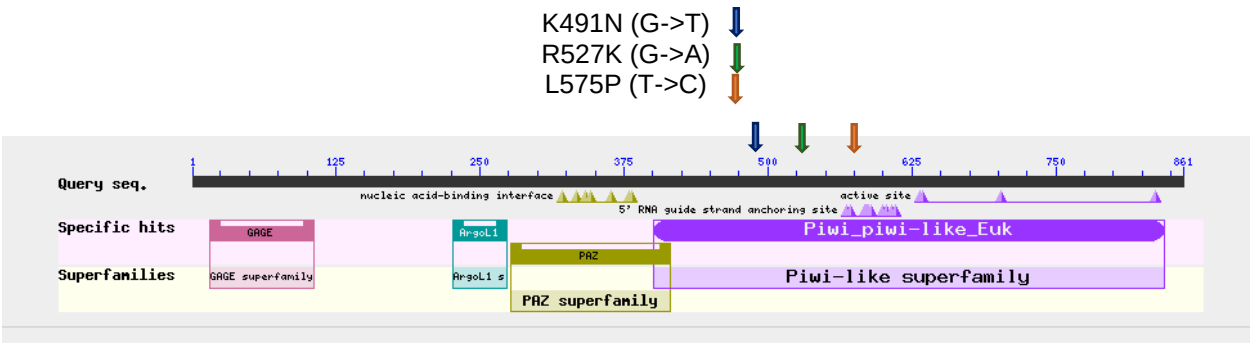
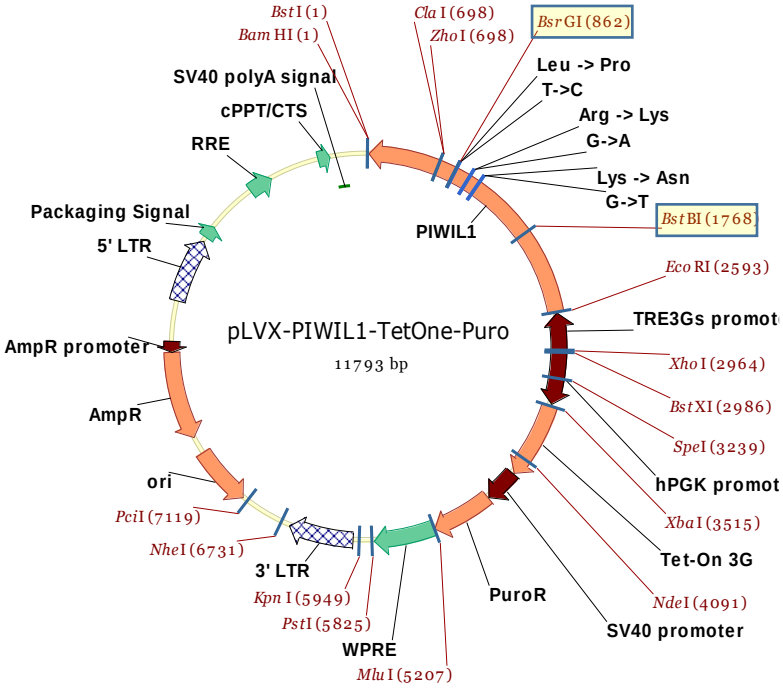


Figure 4. Predicted effect of PIWIL1 variation on protein activity. All three of the variant nucleotides of our induced transcript compared to the new PIWIL1 sequence updated online were located in the PIWI-interacting domain. The predicted effect of the third variant, L575P would be predicted to be especially important as a leucine to a proline change implies that the protein would take on a significant bend.

Figure 5. L575P would likely affect protein structure. Feature 1 in this diagram is predicted to be the active site of PIWIL1's RNAse H activity, based on structural comparisons with Agonaute (as per conserved sequences comparison at the NCBI website). Our sequence is "query". The gene ID numbers of related sequences is listed. None of the other proteins have a proline in this position, again supporting our hypothesis that this amino acid change might affect protein function (see highlighted protein at amino acid 575 for "query").

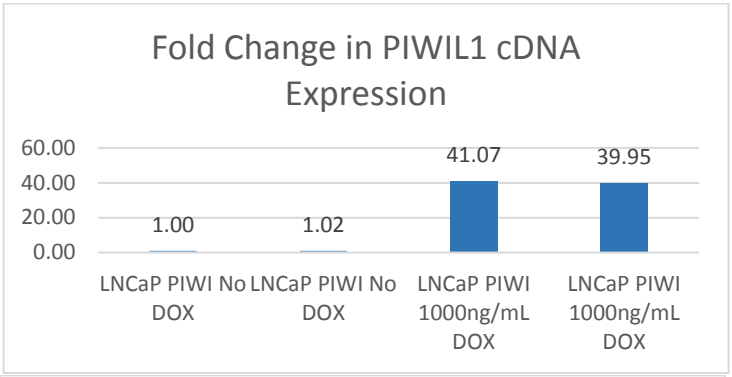
As these changes might affect protein stability, we decided to replace all three with the updated sequence on NCBI's website. All three changes were within a 300 base pair region, so this was relatively painless as we were able to have a synthetic piece of DNA synthesized, and cloned into the first construct, thus we could directly compare the difference the variant amino acids made functionally.

Figure 6. Cloning strategy to replace three variant amino acid bases with the updated PIWIL1 sequence. A904 fragment of PIWIL1 was synthesized to replace the three mutations in the first clone, and subcloned into the BstBI – BsrGI sites of the pLVX-PIWIL1-TetOne-Puro recreating the wild type PIWIL1.



Next, stable cell lines were once again made carrying this vector, pLVX-PIWIL1-TetOne_Puro_new.

a



b

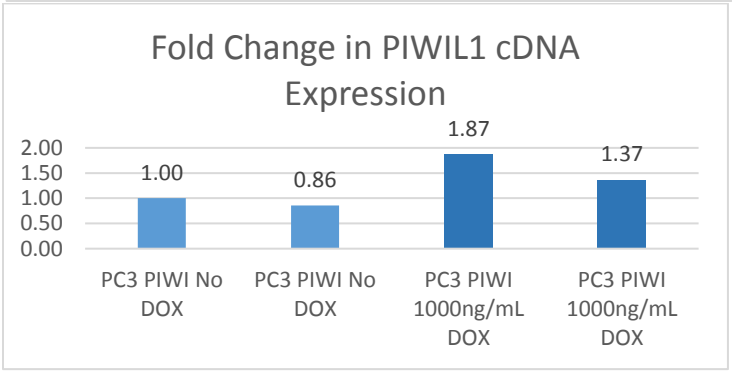
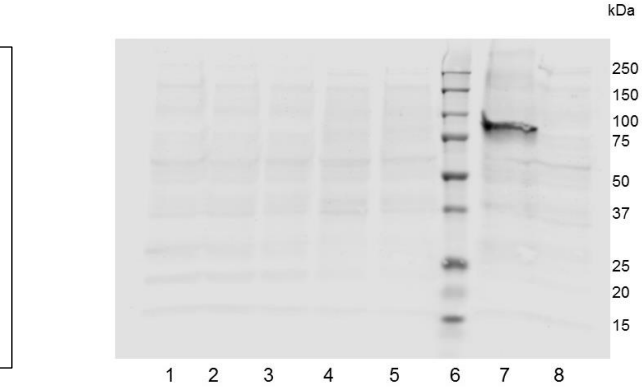


Figure 7. Induction of PIWIL1 transcript from new vector sequence. In part A, we can see that real-time PCR shows induction of PIWI at 40 fold (approximately) with 72 hours treatment of 1000ng/ml doxycycline. In part B, we see that in PC3 cells, which express a small amount of PIWIL1 transcript intrinsically, induction is less than two-fold.

At this point there was no change in cell growth in either cell line, as happened with the

Figure 8. Western blot for PIWIL1 after induction of transcripts in LNCaP and PC3 cells.



We can see that the PIWIL1 protein is detected in Figure 8, but only in the positive control (mouse testis lysate). Our conclusions from these experiments was that the first PIWIL1 sequence we started with had 3 amino acid differences that likely affected the stability of the protein, therefore despite robust induction of transcript expression after producing stably transduced cell lines, the protein was degraded. This is not likely to be due to our handling of the protein, as we routinely use protease inhibitors and these cell

lysates were always rapidly processed on ice to avoid protein degradation. Additionally, we can that our positive control which was from frozen mouse tissue did not degrade. After we produced the second construct, induction was nearly absent in PC3 cells and far reduced relative to the first construct in LNCaP cells, despite utilizing the same conditions for both experiments. We also assessed expression using different concentrations and time points of DOX, with no difference in the results (data not shown). It is my opinion, although not proven yet, that functional PIWIL1 protein expression likely is so toxic to the cell lines that any cells that did express protein rapidly died, possibly due to leakiness of the tet-on promoter driving the constructs. We did use tet-free fetal bovine serum in these experiments and when growing the stable cell lines, however it is possible that there was some transcription of the gene and if it were highly toxic to the cell then those clones would have died early on during selection. It is notable that of all the NCI60 cancer cell lines, the only one with any transcript expression of PIWIL1 is PC3, however clearly that transcript doesn't produce enough protein in itself that is detectable by western blot or else we would have seen its expression in our un-induced controls.

2. LINE-1 expression in hybrid sequences derived from prostate cancer metastases

For this part of the project we planned to assess LINE-1 (L1) activation of adjacent sequences in metastases to the lymph node from prostate cancer. We obtained 3 independent lymph node metastases and also used the LNCaP cell line, which in itself is derived from a lymph node metastasis of the prostate. We used paired-end 150bp RNA sequencing to generate the data, but then we had to use (and actually develop) a novel pipe-line for analyzing this data. This was eventually achieved by working with a bioinformatician and an example of the pipe-line we used is below (Figure 9, next page).

Typical RNA-seq pipelines eliminate hybrid sequences as one of their first steps, and in addition, L1 sequences are "masked" out, as they are part of the repetitive sequence family and therefore difficult to map to the reference genome. Because of this, we designed a pipeline where we first screened sequences for the presence of L1 sequences. The number of reads for each of the 4 samples is listed in Table 1 below.

Table 1: Read statistics

Sample Name	# of reads	[1]# of pairs with 1 read mapped to Human L1	[2]# of singleton reads mapped to Human L1	Mapped to human genome
LNCap	64,451,999	92,141	6,496	4,847
1348A	72,049,017	1,156,021	64,923	49,740
1348F	68,975,740	1,082,085	70,731	55,307
1348H	77,271,251	1,242,113	80,092	62,703

1) Without subtracting secondary alignment
2) Only keep those sequences that one-end mapped to L1, but the other end does not.

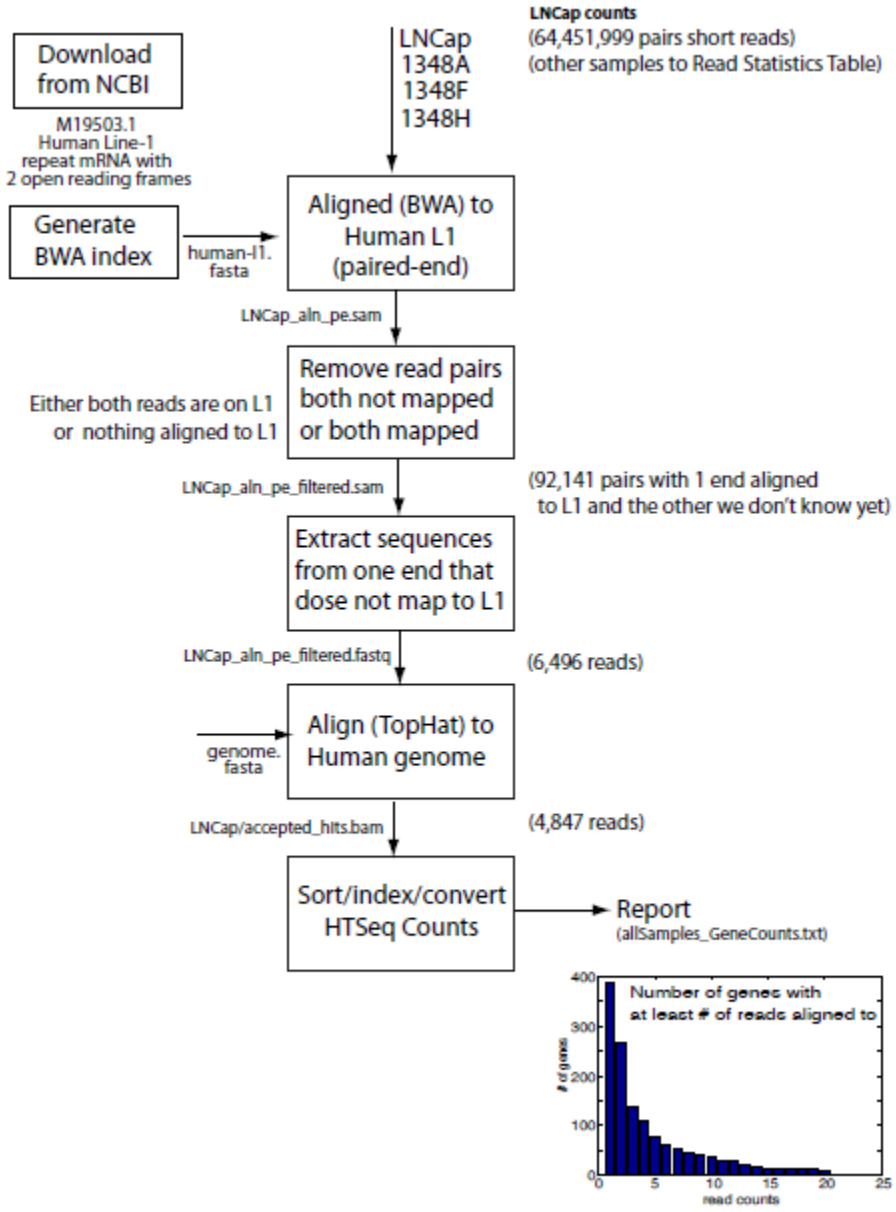
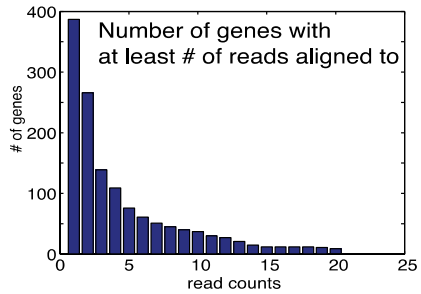


Figure 9. Novel pipeline to identify hybrid LINE-1 (L1) sequences from metastatic to the lymph node prostate cancer. Four independent samples were subjected to paired-end RNA sequencing with 150 base pair read-lengths. This allowed for the best possible chance of identifying fusion sequences, which would not be able to be identified easily with conventional single read sequencing.

As shown in Table 2, using this method we were able to identify 387 activated transcripts among the four samples, of these 37 genes showed 10 or more hits in total over the 4 samples, and Table 3 shows the numbers of genes with reads from L1 from the 4 samples, allowing us to narrow down hybrid transcripts that might be of importance

Table 2: Counts aligned to genes with at least one end on Human L1.



	# of reads aligned to genes (summed from all samples)									
	≥1	≥2	≥3	≥4	≥5	≥6	≥7	≥8	≥9	≥10
# of genes	387	266	139	109	76	61	51	45	40	37

Table 3: Number of genes detected in each sample

# of genes with reads (sum over 4 samples)	LNCaP	1348A	1348F	1348H
≥ 1	57	207	241	278
≥ 5	7	16	17	32
≥ 10	3	5	6	9

Using this analysis method, we generated the below table as the top 3 candidates for further scrutiny.

Table 4: top three gene targets for L1 activation by sample.

	Read counts			
Gene	LNCaP	1348A	1348F	1348H
AAK1	30	46	50	34
LPP	14	48	61	42
WDR72	25	13	35	15

What are these genes and are they known to be linked to cancer?

AAK1 – Adaptor Protein 2 Associated Kinase 1 – According to the literature, this gene regulates clathrin-mediated endocytosis by phosphorylating the AP2M1/mu2 subunit of the adaptor protein complex 2 (AP-2). It is also thought to phosphorylate NUMB to regulate its cellular localization, promoting NUMB localization to endosomes. Finally AAK1 binds to and stabilizes the activated form of NOTCH1, increases its localization in endosomes and regulates its transcriptional activity. As we know, NOTCH1 activation is intimately involved with prostate cancer progression. This transcript was upregulated in all four independent samples suggesting a potential selection for a pro-carcinogenic transcript.

LPP - LIM Domain Containing Preferred Translocation Partner In Lipoma – This gene has been shown to pro-oncogenic when formed as a fusion partner. However very recently this gene has been showed to be an Src substrate that is required for efficient metastasis in breast cancer (PMC5413977). Therefore it may well play a role in prostate cancer metastasis as well, although a pubmed search shows no manuscripts describing this gene in prostate cancer.

WDR72 – WD repeat-containing protein 72 –this protein plays a critical role in calcium transport and matrix protein removal during enamel maturation in teeth (PMC4521966). Other than this, little is known about this protein. Recent literature points to a role of this protein in brain metastasis from thyroid carcinoma (PMC5088282), and esophageal cancer (PMID26631031). Clearly there is a need to study this gene further.

These transcripts were activated in all 4 samples, however there were other transcripts that were just activated in 2 or 3 of the samples, which may also be important. We are currently in the process of banking prostate tissues and metastases at UTHSCA therefore we will be able to assess these genes further in the coming months, although this is the final report for this project.